Follow Your Pose: Pose-Guided Text-to-Video Generation using Pose-Free Videos

Year: 2023 | Citations: 258 | Authors: Yue Ma, Yin-Yin He, Xiaodong Cun, Xintao Wang, Ying Shan

---

Abstract

Generating text-editable and pose-controllable character videos have an imperious demand in creating various digital human. Nevertheless, this task has been restricted by the absence of a comprehensive dataset featuring paired video-pose captions and the generative prior models for videos. In this work, we design a novel two-stage training scheme that can utilize easily obtained datasets (i.e., image pose pair and pose-free video) and the pre-trained text-to-image (T2I) model to obtain the pose-controllable character videos. Specifically, in the first stage, only the keypoint image pairs are used only for a controllable text-to-image generation. We learn a zero-initialized convolutional encoder to encode the pose information. In the second stage, we finetune the motion of the above network via a pose-free video dataset by adding the learnable temporal self-attention and reformed cross-frame self-attention blocks. Powered by our new designs, our method successfully generates continuously pose-controllable character videos while keeps the editing and concept composition ability of the pre-trained T2I model. The code and models are available on https://follow-your-pose.github.io/.